基于极性转移和 LSTM 的树结构网络与句子分类 *

汪 冉，金 忠

(南京理工大学 计算机科学与工程学院，南京 210018)

摘 要： 长短期记忆网络(long short term memory，LSTM)是一种能长久储存序列信息的循环神经网络，在语言模型、语音识别、机器翻译等领域都得到了广泛的应用。先研究了前人如何将 LSTM 中的记忆模块拓展到语法树得到 LSTM 树结构网络模型，以获取和储存句子深层次的语义结构信息；然后针对句子词语间的极性转移在 LSTM 树结构网络模型中添加了极性转移信息提出了极性转移 LSTM 树结构网络模型，更好获取情感信息来进行句子分类。实验表明在 Stanford sentiment tree-bank 数据集上，提出的极性转移 LSTM 树结构网络模型的句子分类效果优于 LSTM、递归神经网络等模型。

关键词： 神经网络；LSTM 网络；树结构网络；极性转移；句子分类

中图分类号： TP393　　doi: 10.3969/j.issn.1001-3695.2017.07.0669

Tree-structured networks based on polarity shifting and LSTM for sentences classification

Wang Ran, Jin Zhong

(*School of Computer Science & Engineering*, *Nanjing University of Science & Technology*, *Nanjing* 210018, *China*)

Abstract: Long short term memory (LSTM) is a recurrent neural network (RNN) which has outstanding ability to preserve sequence information for a long time. It has been widely used in the language modeling, machine translation, speech recognition and other fields. Firstly, this paper explored how predecessors extend the memory module in LSTM to the syntax tree to get Tree-Structured LSTM networks model which can obtain and store the semantic structure information of the sentences; Then, according to the polarity shifting information between the words of sentences, it proposed a Polarity Shifting Tree-Structured LSTM networks model to capture the sentiment information for better sentences classification. The proposed model show a better performance than LSTM, recursive neural networks and other models for sentences classification on Stanford Sentiment Tree-bank dataset.

Key Words: neural networks; LSTM; tree-structure network; polarity shifting; sentence classification

0 引言

近年来深度学习被有效地广泛应用于语音识别[1]、机器翻译[2]以及图像转换成文本等领域[3]。在自然语言处理领域，随着学者对神经网络的不断探索，各种神经网络模型在语言建模和文本情感分类等方面都取得了很大的进展。

句子分类在自然语言处理领域具有很大的研究价值和应用价值[4]，该研究内容引起越来越多的国内外学者的兴趣和重视。使用神经网络句子分类时，首先面对的是句子建模，即用模型训练后的向量来表示句子。目前有三种模型应用比较广泛：词袋模型、序列模型和树结构模型。三种模式各有特点，词袋模型比较简单，它将句子看成无序的单词的集合，有时将所有单词向量的平均向量作为句子的向量表示[5,6]，它完全忽略单词的顺序和语法，很难获取句子有意义的信息。该模型经常被很有效地用在传统的贝叶斯模型、LSA 和 LDA 等模型中。2008 年 Pang 和 Lee 等人使用词袋模型对文本进行情感分析和分类，随后更多人试着设计能获取更好的句子特征的模型或者使用基于句法结构的极性转移特征来提高情感分类的准确率。序列模型则很看重句子中所有单词的排列顺序[7,8]，该模型句子建模时直接将单词向量按排列顺序连接得到句子表示，然而并没有考虑到句子的结构信息。而树结构模型是根据给定的句法树构造递归网络最终获得句子中短语和句子的表示[9,10]。树结构已经用于分析自然的图像场景，还用于从词向量到短语向量的转换，并对句子的情感极性进行分类[11]。2013 年 Socher 等人在影评

基金项目：国家自然科学基金资助项目（61373063，61375007，61233011）

作者简介：汪冉（1993-），女，江苏南京人，硕士研究生，主要研究方向为深度学习、自然语言处理（wangranahu@163.com）；金忠（1961-），男，江苏南京人，教授，博导，主要研究方向为模式识别、图像分析、机器学习等.基金项目：国家自然科学基金项目（批准号 61373063，61375007，61233011）
作者简介：汪冉（1993-），女，南京，硕士研究生，主要研究方向为深度学习，自然语言处理，wangranahu@163.com；金忠（1961-），男，南京，博导（教授），主要研究方向为模式识别，图像分析，机器学习等，zhongjin@njust.edu.cn.

数据上使用递归神经网络模型进行情感分类展现了比传统模型更好的效果。

句子分类中，词袋模型并不能完全掌握自然语言句子的全部语义，无法分辨由词序或句法结构造成的语义差别，于是使用序列模型和树结构模型会更有效，而且由于树结构模型与句法结构的相关性更强，树结构模型是最优选择。循环神经网络（RNN）是解决序列化的一种有效的方法，理论上它可以处理任意长度的序列，但在训练网络时会出现梯度消失的问题[12]。LSTM 网络是 RNN 的一种改进，该网络在处理梯度消失问题上有很大突破[13]，使网络可以长久保存序列信息，但 LSTM 网络还是无法获得句子的句法结构信息。在基于树的递归神经网络（recursive neural network, RcNN）结构中运用 LSTM 记忆模块的 Tree-Structured LSTM 模型则不仅能够获取句法结构和语义信息还能在递归中长久保存节点信息。原始的 LSTM 记忆模块输入是当前时刻的外部输入以及上一时刻的隐藏状态量，Tree-Structured LSTM 模型的输入则是当前节点的外部输入和子节点的隐藏状态量。

本文先介绍了已有的基本网络模型，重点介绍将 LSTM 记忆模块分别运用到基于依存句法树和短语结构树的树结构网络中得到的 tree-structured LSTM 模型；然后将情感极性转移融合到 tree-structured LSTM 模型中提出了两种新的 *polarity tree-structured LSTM (PTree-structured* LSTM)模型，丰富了 LSTM 记忆模块的信息，不仅获得句子的语义信息和句法结构信息还保存了情感信息，更有效地解决句子分类问题。接着给出了在本文模型下实验的结果以及和其他模型的对比分析。

1　相关工作

1.1　长短期记忆网络（LSTM）

循环神经网络（RNN）通过对网络块中隐藏状态向量 $h_t \in \Re^d$ 的循环使用，可以处理任意长度的序列。其中，t 时刻的隐藏状态向量 h_t 是由此刻的输入向量 x_t 和 $t-1$ 时刻的隐藏状态向量 h_{t-1} 的非线性函数转换得到，如式(1)所示的双曲正切函数：

$$h_t = \tanh(Wx_t + Uh_{t-1} + b) \tag{1}$$

但是，RNN 在训练网络参数时会产生梯度向量下降或者增长速度过快的问题[15,16]，特别处理一个较长序列时，RNN 模型会因为存在此问题而很难有很好的效果。

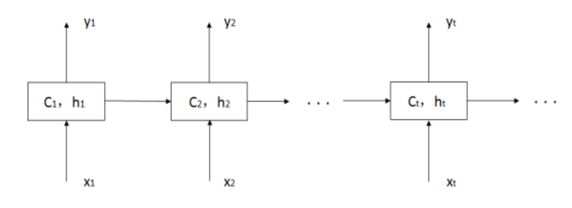


图 1　LSTM 网络结构

如图 1 所示，长短期记忆（LSTM）网络通过利用一个记忆模块来改进 RNN，它在此模块中保存若干状态向量来保存序列中的每个时刻信息。这些状态向量包含：输入门 i_t，遗忘门 f_t，输出门 o_t，记忆状态 c_t，以及隐藏状态 h_t。LSTM 记忆模块中的向量维数 d 称为 LSTM 的内部向量维度。LSTM 单元中的状态向量转换公式如下：

$$i_t = \sigma(W^{(i)}x_t + U^{(i)}h_{t-1} + b^{(i)}) \tag{2}$$

$$f_t = \sigma(W^{(f)}x_t + U^{(f)}h_{t-1} + b^{(f)}) \tag{3}$$

$$o_t = \sigma(W^{(o)}x_t + U^{(o)}h_{t-1} + b^{(o)}) \tag{4}$$

$$u_t = \tanh(W^{(u)}x_t + U^{(u)}h_{t-1} + b^{(u)}) \tag{5}$$

$$c_t = i_t \odot u_t + f_t \odot c_{t-1} \tag{6}$$

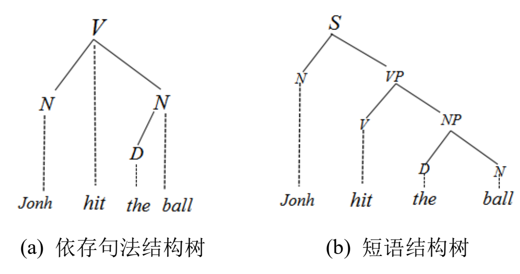
$$h_t = o_t \odot \tanh(c_t) \tag{7}$$

在式(2)~(7)中，x_t 是时刻 t 的输入，Sigmoid 函数将状态向量的数值控制在 0 与 1 之间。该模型通过使用各种门限改变每个向量的状态可以处理各种时间尺度的问题。

1.2　基于 LSTM 的树结构网络

句子中单词之间不仅存在顺序关系，基本的 LSTM 网络局限于序列的有序性不能展示句子的结构信息。而树结构网络可利用句子构造的语法树来建造递归网络，此递归网络可以体现句子结构信息却易造成信息丢失。 2015 年 Socher 等人结合两者的特点将 LSTM 记忆模块拓展运用到树结构网络中，提出两种融合 LSTM 记忆模块的树结构网络[17]：融合 LSTM 的依存句法树结构网络 Dependency Tree LSTM（DTree-LSTM）、融合 LSTM 的短语结构树网络 Constituency Tree LSTM（CTree-LSTM）。这种模型近期用于句子表示[18]，句子分析[19]都有很好的效果。

Tree-structured LSTM 中节点的 LSTM 记忆模块中的状态向量均与子节点相关，节点 j 中状态向量包含了输入门 i_j，输出门 o_j，记忆状态 c_j 和隐藏状态 h_j，x_j 表示句子中单词向量，且对每一个子节点均设置一个遗忘门 f_{jk}（K 为孩子节点索引），侧重保存重要的语义信息。



(a) 依存句法结构树　　(b) 短语结构树

图 2　依存句法结构树和短语结构树

1.2.1 融合 LSTM 的依存句法结构树网络

法国的语言学家 L.Tesniere 最早提出依存句法结构,如图 2(a)所示的一棵依存句法树,它描述节点与其子节点之间的依存关系,展示了在句法上词语之间的和语义有关的搭配关系。该树结构有如下特点：子节点是无序的；子节点的数目不定。根据这两个特点，将融合 LSTM 的依存句法树结构（DTree-LSTM）中的状态向量转换设置如下：

$$\tilde{h}_j = \sum_{k \in C(j)} h_k \tag{8}$$

$$i_j = \sigma(W^{(i)}x_j + U^{(i)}\tilde{h}_j + b^{(i)}) \quad (9)$$

$$f_{jk} = \sigma(W^{(f)}x_j + U^{(f)}h_k + b^{(f)}) \quad (10)$$

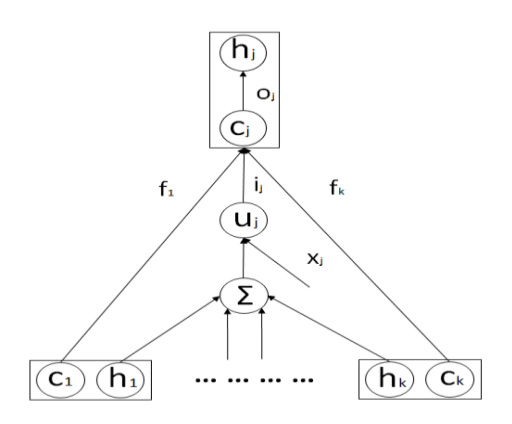
$$o_j = \sigma(W^{(o)}x_j + U^{(o)}\tilde{h}_j + b^{(o)}) \quad (11)$$

$$u_j = \tanh(W^{(u)}x_j + U^{(u)}\tilde{h}_j + b^{(u)}) \quad (12)$$

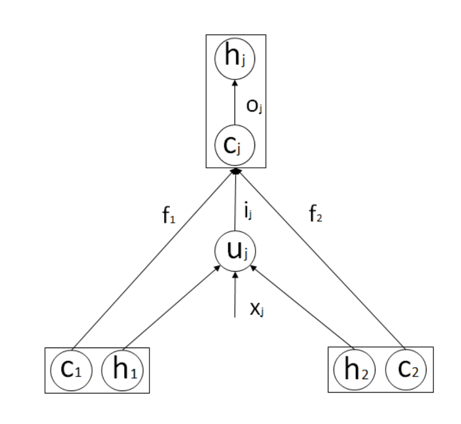
$$c_j = i_j \odot u_j + \sum_{k\in C(j)} f_{jk} \odot c_k \quad (13)$$

$$h_j = o_j \odot \tanh(c_j) \quad (14)$$

式(8)~(14)中的 $C(j)$ 为节点 j 的子节点数目，参数矩阵 W 和 U 将节点的 x_j 和 h_k 转成记忆模块内部向量。DTree-LSTM 网络结构如图 3(a)所示。



(a)DTree-LSTM 结构



（b）CTree-LSTM 结构

图 3　DTree-LSTM 网络结构和 CTree-LSTM 结构

1.2.2 融合 LSTM 的短语结构树网络

句子的短语结构句法分析源于传统的句子图解法，即如图 2(b)所示把句子分割为短语（动词词组，名词词组等）。这种树的非叶子节点表示子节点之间的关系，叶子节点则代表句子中的单词。短语结构树有如下特点：它是一个二叉树，且节点的子节点是有序的；它只有在叶子节点上有单词输入。由其特点，将 LSTM 单元融合到短语结构树（CTree-LSTM）中的状态向量转换设置如下：

$$i_j = \sigma(W^{(i)}x_j + \sum_{l=1}^{N} U_l^{(i)}h_{jl} + b^{(i)}) \quad (15)$$

$$f_{jk} = \sigma(W^{(f)}x_j + \sum_{l=1}^{N} U_{kl}^{(f)}h_{jl} + b^{(f)}) \quad (16)$$

$$o_j = \sigma(W^{(0)}x_j + \sum_{l=1}^{N} U_l^{(o)}h_{jl} + b^{(o)}) \quad (17)$$

$$u_j = \tanh(W^{(u)}x_j + \sum_{l=1}^{N} U_l^{(u)}h_{jl} + b^{(u)}) \quad (18)$$

$$c_j = i_j \odot u_j + \sum_{l=1}^{N} f_{jl} \odot c_{jl} \quad (19)$$

$$h_j = o_j \odot \tanh(c_j) \quad (20)$$

式(15)~(20)中的 N 为节点的子节点数目，$k = 1,2,\ldots,N$。在短语结构树中 N=2 或者 0。该树网络对子节点的隐藏状态 h_k 赋予了不同的参数矩阵。这种做法对短语组合有不同的侧重度，对于情感强烈的节点参数矩阵需调整到更大值。CTree-LSTM 网络如图 3(b)所示。

2　融合极性转移和 LSTM 的树结构网络

2.1　情感极性转移

语言理解中，单词的极性转移能够影响整个句子的情感极性，句子中的副词，否定词等都会使整个句子的极性发生偏转，情感极性转移如表 1 所示一般分为四类：三种句内显式（强调、否定和转折），一种句间极性转折[20]。在 tree-structured LSTM 中，忽略了每个节点的情感信息，所以提出了 PTree-LSTM 网络，在每个节点利用 SoftMax 得出节点的情感状态向量，将这些情感状态向量作为情感极性转移特征更好地训练网络。则在计算节点的表示时，不仅拥有子节点由原始记忆模块保存的语义结构相关信息，还获得了子节点的情感信息，具体的做法是为每一个节点的 LSTM 记忆模块的状态向量中设置添加情感状态向量 l 和情感极性转移向量 b。

表 1　句子中单词极性转移示例

类型	样例	线索	作用域
强调	他很喜欢这本书	显式	句内
否定	我不喜欢这个地方	显式	句内
转折	这个手机还不错，但我不喜欢	显式	句内
句间极性转折	这个手机配置不错，但是样式不好看	隐式	句间

2.2　融合极性转移和 LSTM 的依存句法树网络

在依存句法树中，子节点是无序的，所以对于子节点的情感状态向量 l_k 无须侧重考虑，则状态向量转换设置如式(21)~(30)所示。

$$\tilde{l} = \sum_{k\in C(j)} l_k \quad (21)$$

$$b_j = \sigma(W^{(b)}\tilde{l} + b^{(b)}) \quad (22)$$

$$\tilde{h}_j = \sum_{k \in C(j)} h_k \tag{23}$$

$$i_j = \sigma(W^{(i)} x_j + U^{(i)} \tilde{h}_j + b_j) \tag{24}$$

$$f_{jk} = \sigma(W^{(f)} x_j + U^{(f)} h_k + b_j) \tag{25}$$

$$o_j = \sigma(W^{(o)} x_j + U^{(o)} \tilde{h}_j + b_j) \tag{26}$$

$$u_j = \tanh(W^{(u)} x_j + U^{(u)} \tilde{h}_j + b_j) \tag{27}$$

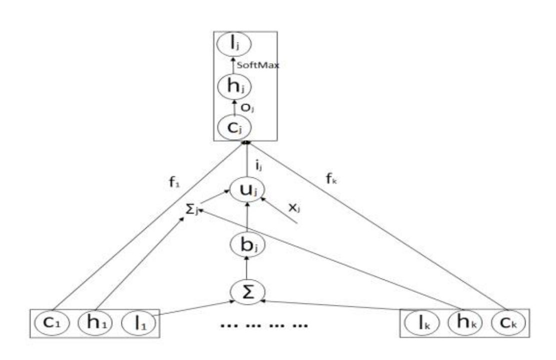
$$c_j = i_j \odot u_j + \sum_{k \in C(j)} f_{jk} \odot c_k \tag{28}$$

$$h_j = o_j \odot \tanh(c_j) \tag{29}$$

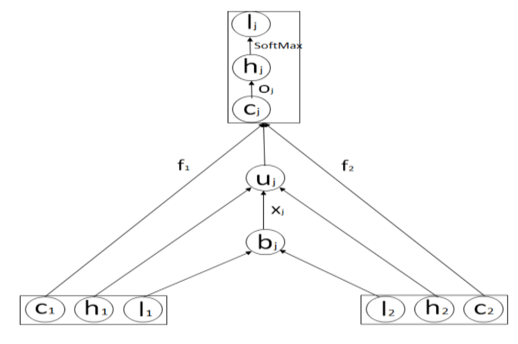
$$l_j = SoftMax(W^{(s)} h_j + b^{(s)}) \tag{30}$$

x_j 表示句子中单词向量，节点中门限向量 i_t，f_t 和 o_t 中的向量值均为 0 到 1 之间。遗忘门 f_t 控制着子节点的记忆模块中哪些内容将被丢弃，输入门 i_t 控制着更新内容，输出门 o_t 控制着输出内容。若在一棵依存句法树中，某节点语义信息能够强烈表达情感，训练网络过程中则会不断调整 $W^{(i)}$ 使 i_j 中的值更接近 1 来保存这个节点的信息，反之则使其更接近 0。

节点的情感极性转移向量 b_j 由综合考虑子节点的情感状态向量 l_k 所得，b_j 表示语法树中单词之间的极性转移信息。网络训练时通过不断调整参数 $W^{(b)}$，由下往上的递归传递极性转移信息来更准确的获得每个节点的情感标签和整个句子的情感信息。 PDTree-LSTM 网络图如图 4(a)所示。



(a)PDTree-LSTM 结构



(b)PCTree-LSTM 结构

图 4　PDTree-LSTM 网络图和 PCTree-LSTM 结构

2.3　融合极性转移和 LSTM 的短语结构树网络

短语结构树中子节点是短语组合，两个节点的侧重性应该不同，所以极性转移在此模型中更有突出体现，则需将子节点的情感状态向量分开考虑，状态向量转换设置如式(31)~(38)所示。

$$b_j = \sigma(\sum_{l=1}^{N} W_l^{(b)} l_{jl} + b^{(b)}) \tag{31}$$

$$i_j = \sigma(W^{(i)} x_j + \sum_{l=1}^{N} U_l^{(i)} h_{jl} + b_j) \tag{32}$$

$$f_{jk} = \sigma(W^{(f)} x_j + \sum_{l=1}^{N} U_{kl}^{(f)} h_{jl} + b_j) \tag{33}$$

$$o_j = \sigma(W^{(0)} x_j + \sum_{l=1}^{N} U_l^{(o)} h_{jl} + b_j) \tag{34}$$

$$u_j = \tanh(W^{(u)} x_j + \sum_{l=1}^{N} U_l^{(u)} h_{jl} + b_j) \tag{35}$$

$$c_j = i_j \odot u_j + \sum_{l=1}^{N} f_{jl} \odot c_{jl} \tag{36}$$

$$h_j = o_j \odot \tanh(c_j) \tag{37}$$

$$l_j = SoftMax(W^{(s)} h_j + b^{(s)}) \tag{38}$$

短语结构树中各个节点表示短语组合，而动词词组的情感表达要远远强烈于名词词组，所以并不能同于依存句法树中将子节点的情感状态向量同等考虑，于是在记忆模块中对 l_{jl} 和 h_{jl} 均给予不同的参数矩阵，且网络训练中会不断调整 $W_l^{(b)}$ 使网络更能侧重保存动词词组的结构信息以及情感信息。由于短语结构树网络对子节点更多的限制，网络中的参数要远多于依存句法树网络中的参数，相同地，该模型在训练网络中通过不断调整 $W_l^{(b)}$ 使节点标签更贴近真实标签。PCTree-LSTM 网络图如图 4(b)所示。

2.4　树结构用于句子分类的误差计算

本节主要介绍如何将提出的模型应用到句子分类中。树结构网络中大部分节点都存在真实标签，每个节点表示以这个节点为根节点的子树的短语。对于任意节点 j，设该节点表示的短语为 $\{x\}_j$，可利用式(39)(40)得到预测标签 $\hat{y}_j$。

$$\hat{p}_\theta(y \mid \{x\}_j) = SoftMax(W^{(s)} h_j + b^{(s)}) \tag{39}$$

$$\hat{y}_j = \arg\max_y \hat{p}_\theta(y \mid \{x\}_j) \tag{40}$$

模型中每个节点的预测标签和真实标签的误差和为整个网络的误差，选用负对数似然函数作为误差计算函数。设树结构有 m 个节点，每个节点的误差计算公式如式(41)（仅计算有真实标签的节点）所示，其中 λ 和 θ 是正则化的超参数。

$$J(\theta) = -\frac{1}{m} \sum_{k=1}^{m} \log \hat{p}_\theta(y^{(k)} \mid \{x\}^{(k)}) + \frac{\lambda}{2} \|\theta\|_2^2 \tag{41}$$

3　实验

3.1　句子分类实验

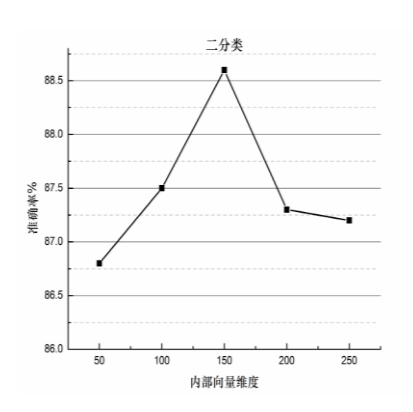
为了测验模型的有效性，使用斯坦福大学基于影评数据创建的 Stanford sentiment tree-bank[21]来进行实验，该数据集有两种类型：用于二分类的句子和用于五分类的句子，五分类为：

very negative, negative, neutral, positive, very positive. 数据集中共有 11855 个句子，平均句子长度为 19，数据集中共有 215154 个短语，数据集提供了短语和句子标签。如果用其做二分类实验，数据集中 6920 句是训练集，872 句是验证集，1821 句是测试集（原标为中性的句子全部去除）；如果用其做五分类实验，数据集中 8544 句式训练集，1101 句是验证集，2210 句是测试集。数据集中已经包含每个句子的短语结构树，本文使用 Stanford 语义分析包获取句子依存句法结构树。

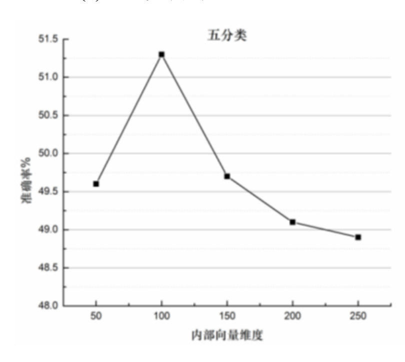
实验中，将 PDTree-LSTM 模型和 PCTree-LSTM 模型都分别进行了二分类和五分类的实验。

3.2 超参数调整

实验中使用 300 维的 Glove vectors 作为单词向量，训练网络过程中词向量的学习率是 0.1，模型使用 AdaGrad 来训练网络，其中学习率是 0.05，批量数值为 25。在整个模型训练中为避免过拟合使用丢弃率为 0.5 的 Dropout。实验发现记忆模块中内部向量维度 d 的选择会影响实验性能，d 较小则使向量保存信息较少，d 较大则会导致参数过多，训练结果易产生过拟合现象。如图 5 所示，由实验得出，二分类中 d 设为 100 最佳，五分类 d 设为 150 最佳。



(a)二分类内部向量维度选择

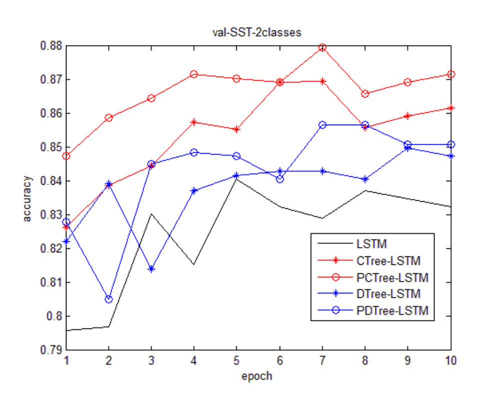


(b)五分类内部向量维度选择

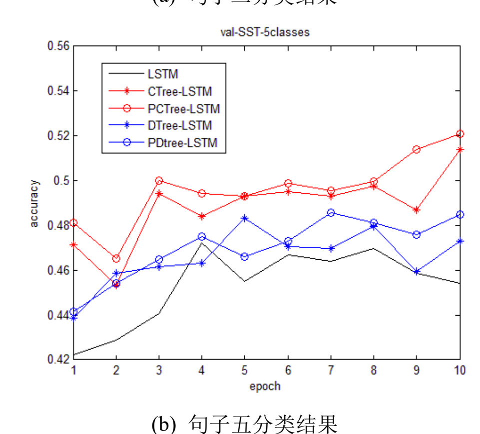
图 5 二分类和五分类内部向量维度选择

3.3 实验结果

实验中 LSTM, CTree-LSTM, PCTree-LSTM, DTree-LSTM, PDTree-LSTM 五个模型在验证集上的测试结果如图 6 所示，本文提出的模型 PCTree-LSTM 和 PDTree-LSTM 在训练过程中与基本的网络模型相比都有很好的表现。



(a) 句子二分类结果



(b) 句子五分类结果

图 6 句子二分类和五分类结果

最后本文模型与各种网络模型在测试集的测试结果对比如表 2 所示，由结果可以看出提出的两个模型在二分类和五分类实验中都能达到很好的效果，其中 PCTree-LSTM 模型在这个数据集上的句子分类效果最好，优于 PDTree-LSTM 模型。这两种模型的实验结果差距可能是来自两种树结构的节点数目的差别。而且短语结构树是以短语关系为节点的树，拥有 31.9 万的存在短语标签的节点，在训练中短语结构树可以获取更多的短语信息，而依存句法树是以结构关系为节点的树，某些节点只能表示为单词之间的关系，其子树单词分布并不能构成一个短语，只拥有 15 万存在短语标签的节点。而且在短语结构树中，参数要远远多于依存句法树，参数越多可以理解为保存的信息越多，这也有影响了分类效果。

表 2 模型分类正确率对比结果

模型	二分类正确率（%）	五分类正确率（%）
DCNN	86.8	48.5
CNN-nonstatic	87.2	48.0
CNN-multichannel	88.1	47.4
LSTM	84.9	47.8
DTree-LSTM	85.7	48.4
PDTree-LSTM	86.2	48.7
CTree-LSTM	88.0	50.9
PCTree-LSTM	**88.6**	**51.3**

4　结束语

本文先介绍了几种已有的十分有效的网络，特别是将 LSTM 记忆模块拓展到树结构网络中，使网络能够有效获取句子的语义结构信息，丰富了网络对句子特征的学习；然后在此树结构网络的 LSTM 记忆模块中添加了情感极性转移信息提出了新的网络模型，进一步加强了模型对句子情感信息的学习。神经网络模型在句子处理上有很大的优势，下一步的工作展望是如何利用神经网络更加快速学习句子文本的结构特征，从而更准确地获取句子情感信息。

参考文献：

[1] Zhang Y, Chan W, Jaitly N. Very deep convolutional networks for end-to-end speech recognition [C]// Proc of IEEE International Conference on Acoustics, Speech and Signal Processing. 2017: 4845-4849.

[2] Gulcehre C, Firat O, Xu K, et al. On integrating a language model into neural machine translation [J]. Computer Speech and Language, 2017, 45 (1): 137-148.

[3] Ma L, Lu Z, Li H. Learning to answer questions from image using convolutional neural network [C]// Proc of the 30th AAAI Conference on Artificial Intelligence. [S. l.] : AAAI Press, 2016: 3567-3573.

[4] 赵妍妍，秦兵，刘挺．文本情感分析 [J]．软件学报, 2010, 21 (8): 1834-1848.

[5] Landauer T K, Dumais S T. A solution to plato's problem: the latent semantic analysis theory of acquisition, induction, and representation of knowledge [J]. Psychological Review, 1997, 104 (2): 211-240.

[6] Foltz P W, Kintsch W, Landauer T K. The measurement of textual coherence with latent semantic analysis [J]. Discourse Processes, 1998, 25 (2-3): 285-307.

[7] Elman J L. Finding structure in time [J]. Cognitive Science, 1990, 14 (2): 179-211.

[8] Mikolov T. Statistical Language models based on neural networks [J]. Presentation at Google, Mountain View, 2012.

[9] Goller C, Kuchler A. Learning task-dependent distributed representations by backpropagation through structure [C]// Proc of IEEE International Conference on Neural Networks, 1996: 347-352.

[10] Socher R, Lin C C, Manning C, et al. Parsing natural scenes and natural language with recursive neural networks [C]// Proc of the 28th International Conference on Machine Learning. 2011: 129-136.

[11] Socher R, Huval B, Manning C D, et al. Semantic compositionality through recursive matrix-vector spaces [C]// Proc of Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning. 2012: 1201-1211.

[12] Socher R, Pennington J, Huang E H, et al. Semi-supervised recursive autoencoders for predicting sentiment distributions [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2011: 151-161.

[13] Hochreiter S, Schmidhuber J. Long short-term memory [J]. Neural Computation, 1997, 9 (8): 1735-1780.

[14] Jordan M I. Serial order: A parallel distributed processing approach [J]. Advances in Psychology, 1997, 121 (2): 471-495.

[15] Hochreiter S. The vanishing gradient problem during learning recurrent neural nets and problem solutions [J]. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 1998, 6 (2): 107-116.

[16] Bengio Y, Simard P, Frasconi P. Learning long-term dependencies with gradient descent is difficult [J]. IEEE Trans on Neural Networks, 1994, 5 (2): 157-166.

[17] Tai K S, Socher R, Manning C D. Improved semantic representations from tree-structured long short-term memory networks [C]// Proc of Association for Computational Linguistics. 2015.

[18] Rath T. Word and Relation Embedding for Sentence Representation [D]. Phoenix: Arizona State University, 2017.

[19] Goldberg Y. Neural Network Methods for Natural Language Processing [J]. Synthesis Lectures on Human Language Technologies, 2017, 10 (1): 1-309.

[20] 张小倩．情感极性转移现象研究及应用 [D]．苏州：苏州大学, 2012.

[21] Socher R, Perelygin A, Wu J Y, et al. Recursive deep models for semantic compositionality over a sentiment treebank [C]// Proc of Conference on Empirical Methods in Natural Language Processing. 2013: 1631-1642.